ROLE OF SOCIAL MEDIA IN US ELECTIONS

Amitabh S Virk

TABLE OF CONTENTS

CONTEXT

RISE OF SOCIAL MEDIA

“What hath god wrought”

An expression of wonder and marvel at something.

- **Social apps like whatsapp, facebook, instagram and twitter**
- There are **330 million monthly** active users and **145 million daily users** with a total of **1.3 billion accounts** in total. 22% of Americans are believed to be on twitter.
- **83% of the total world’s politicians** are on twitter and **two of the top three influential accounts are also political leader’s accounts.**
- **Social Media definition of Free Speech**

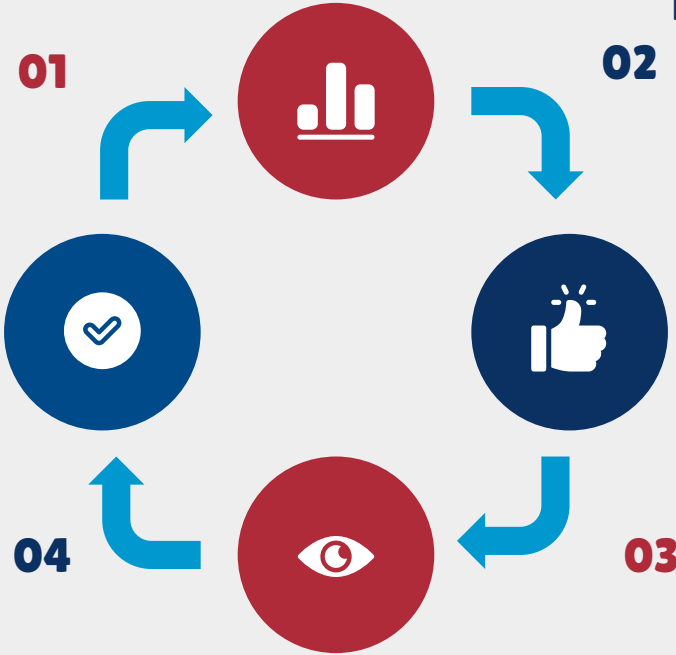
Objective of the Analysis

- Which candidate is more famous on twitter.
- Clusters of tweets based on polarity of the tweets. Negative, Positive and Neutral.
- Locations with most profound sentiments towards candidates.
- What is the probability of negative and positive tweets towards each candidate and locations if possible.
- Prediction of election results based on our analysis.

PROCESS FOLLOWED

01 PROBLEM STATEMENT & RELEVANT DATA COLLECTION

02 PREPROCESSING DATA & CLUSTERING



03 SENTIMENT ANALYSIS

04 SUMMARY & BUSINESS IMPACT

Relevant Data Collection

Tweets

Data is downloaded from kaggle and this data organises tweets based on #Trump and #Biden

Vaccination Data

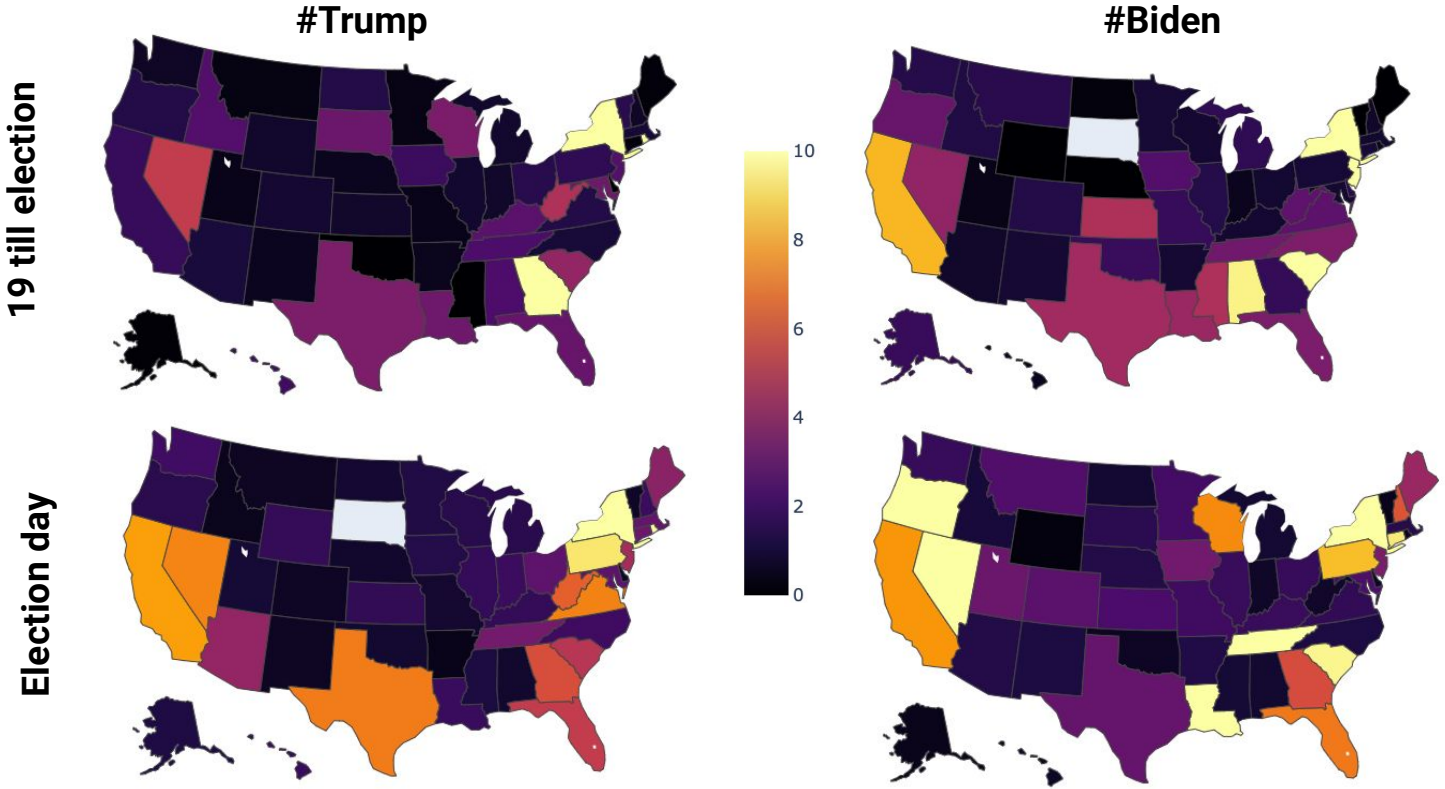
This data is used to see if there is any relation between tweets for particular candidate and Vaccination rates

Location Data

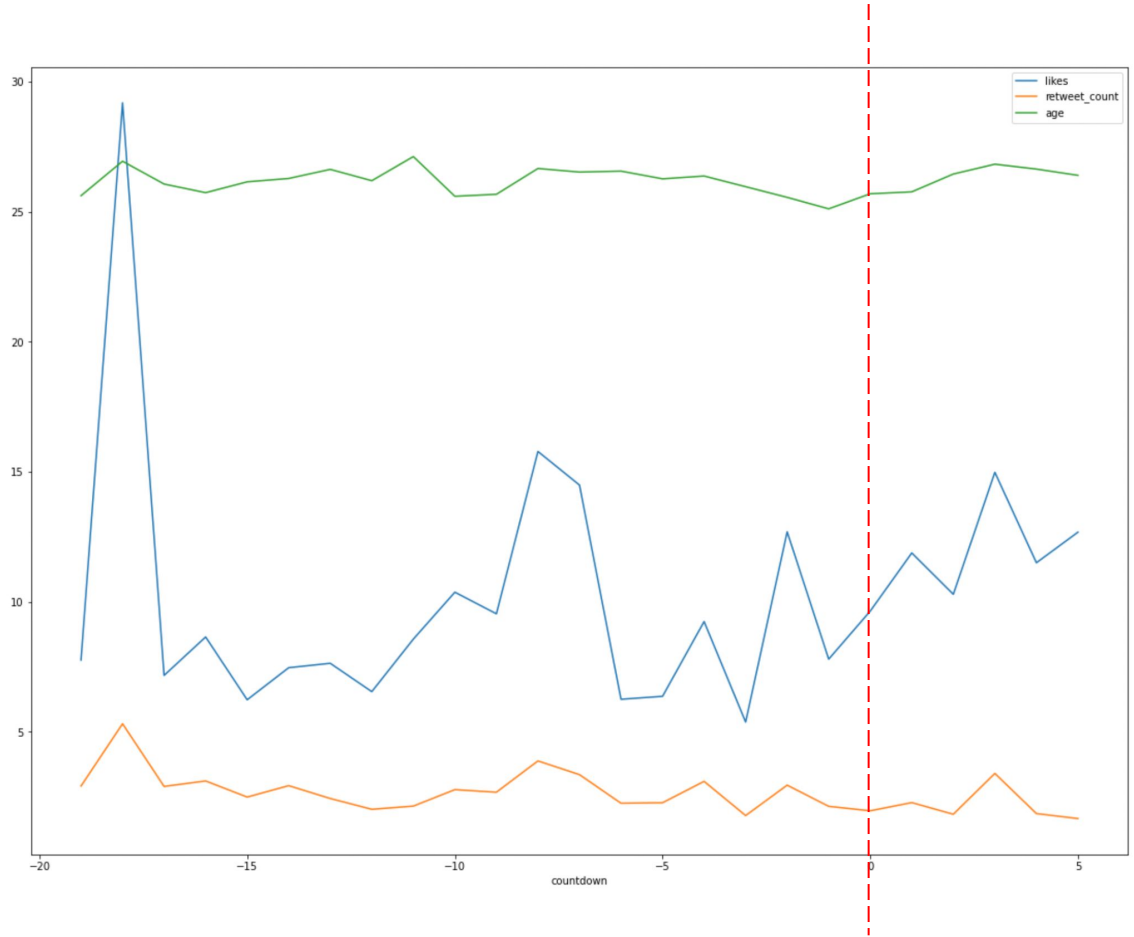
Derived data set from longitude and latitude information embedded in tweets metadata

ANALYSIS

Likes per tweet based on Location



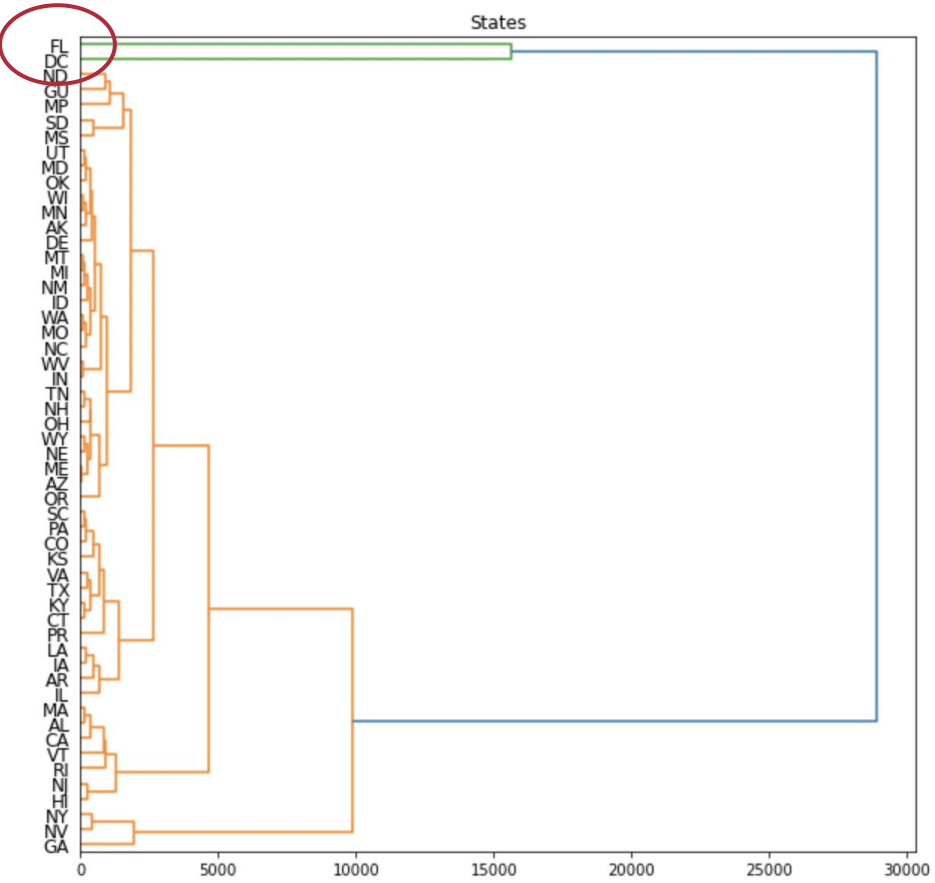
We can see the clear trend that states like **California which are democratic produce Biden tweets with more likes** compared to states like **Texas which are republican and produce more well-liked Trump tweets.**



Trends In the Tweets during election

The following graph shows the averages for likes, retweets, and age of account per day before and after the election.

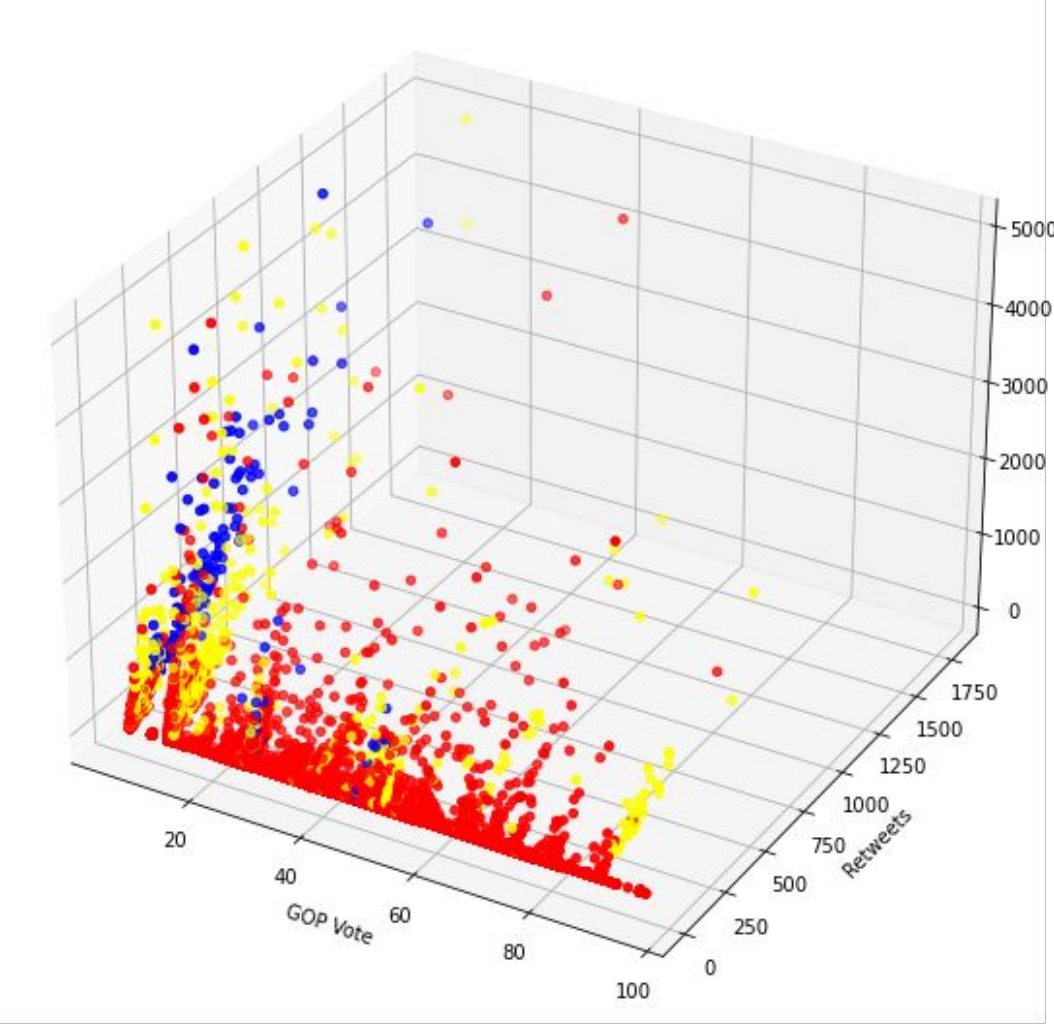
Unexpectedly, there was no spike in new accounts by bots just before the election.



State Clusters

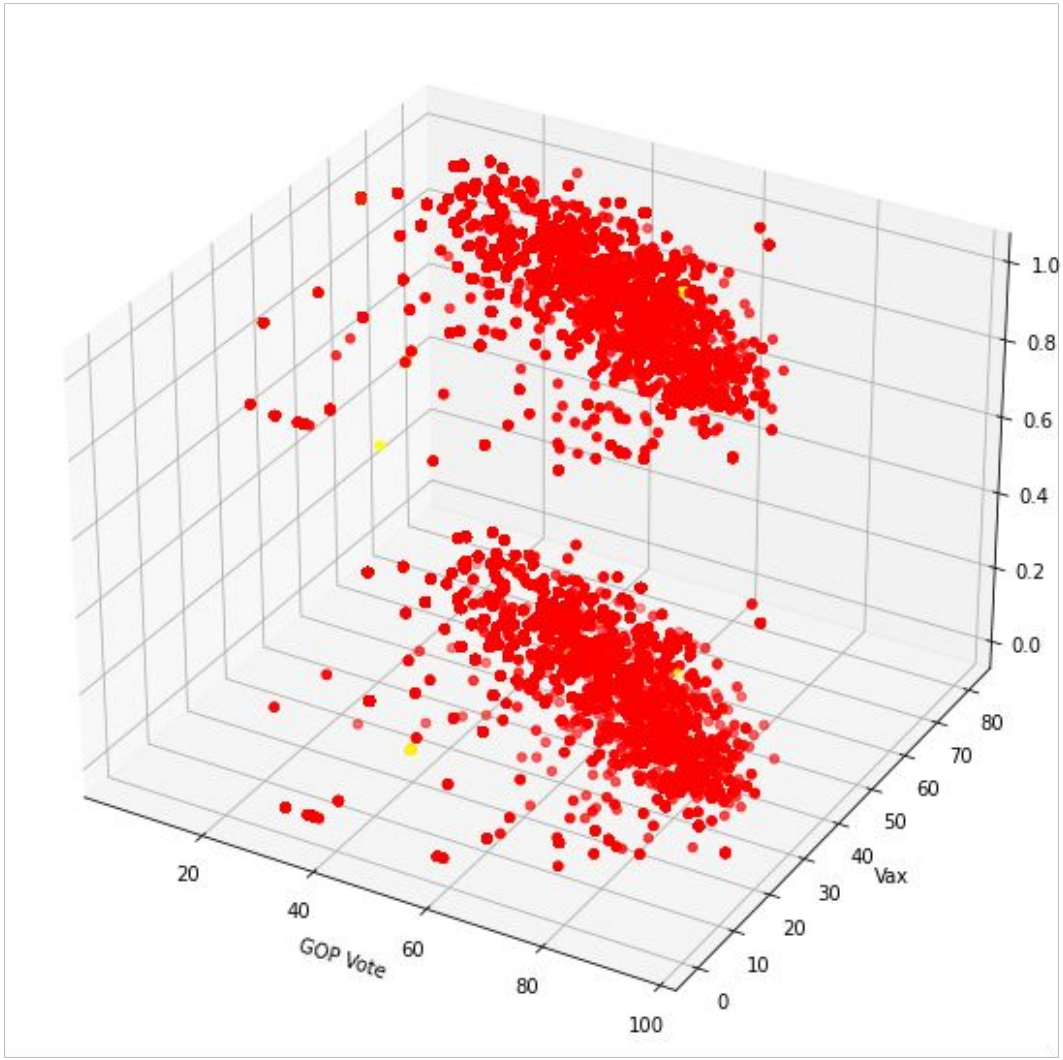
Florida and Washington DC are quite far from the other states. This seems to be driven largely by the huge average follower numbers for DC and Florida accounts

state_code	likes	retweet_count	user_followers_count	countdown	age
DC	53.768958	13.129813	40876.581896	-4.394165	2812.385109
FL	3.765860	1.310973	25254.501214	-5.621946	2624.569657
NY	35.814012	7.160325	14128.377329	-4.813247	2773.118530
NV	9.610371	1.824328	14105.782651	-4.869154	2347.321299
GA	6.084378	1.564079	12237.908887	-4.185033	2840.397007



Location vs Popularity

Blue cluster: From very Democratic counties with higher likes and retweets

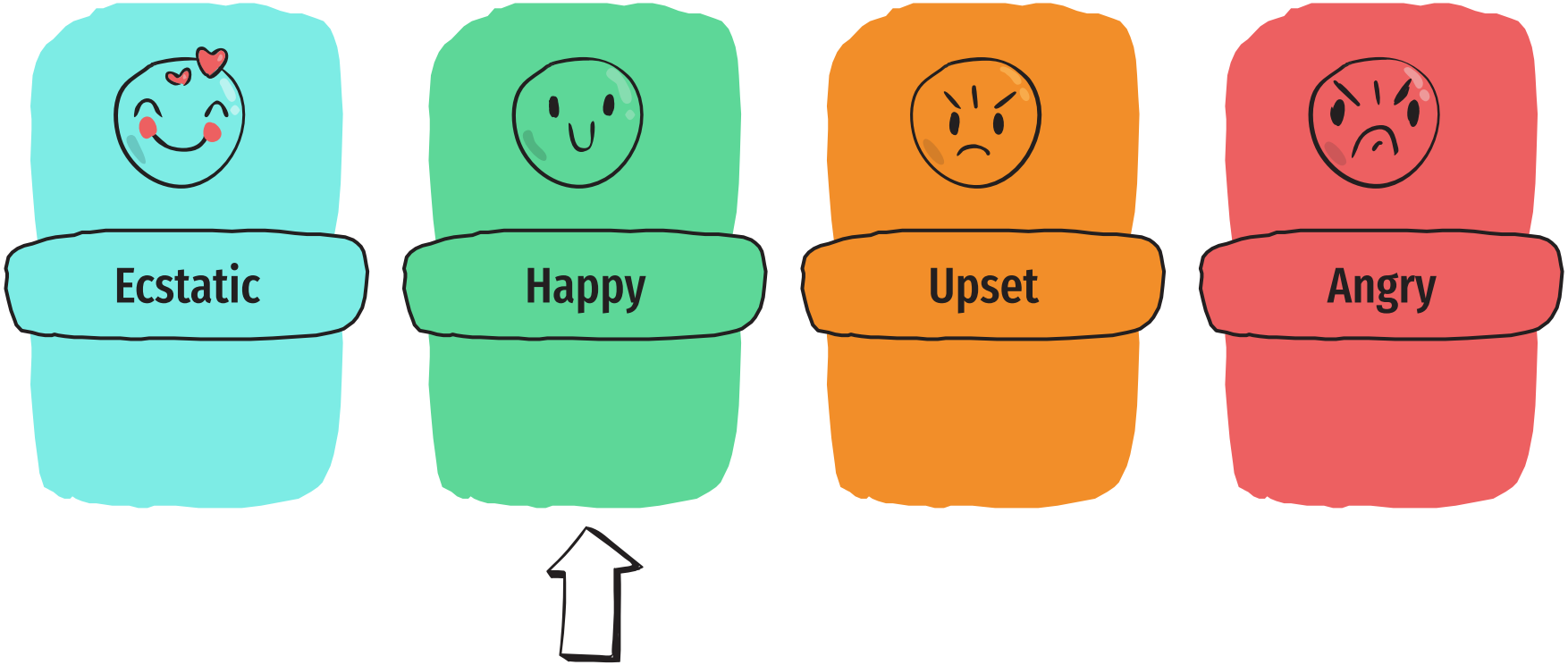


Hashtag vs Politics

Hashtag does not equate with politics:

Even distribution across % GOP Vote and % Vaccination features

Sentiment Analysis



STRATEGY

Step 1
Data Collection

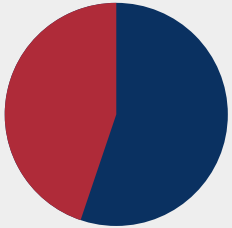
Step 2
Pre-Processing

Step 3
Training Model (Naive Bayes)

Step 4
Testing

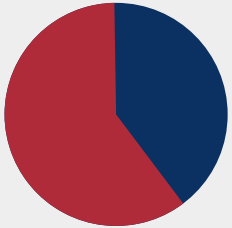
Step 5
Adjusting Hyper Parameters and repeating

Conditional Probabilities?



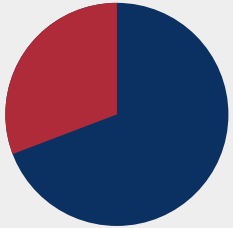
60%

P(+ve |Trump)



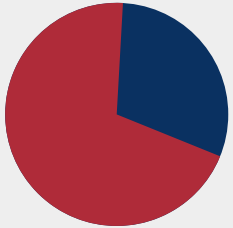
40%

P(-ve |Trump)



70%

P(+ve |Biden)

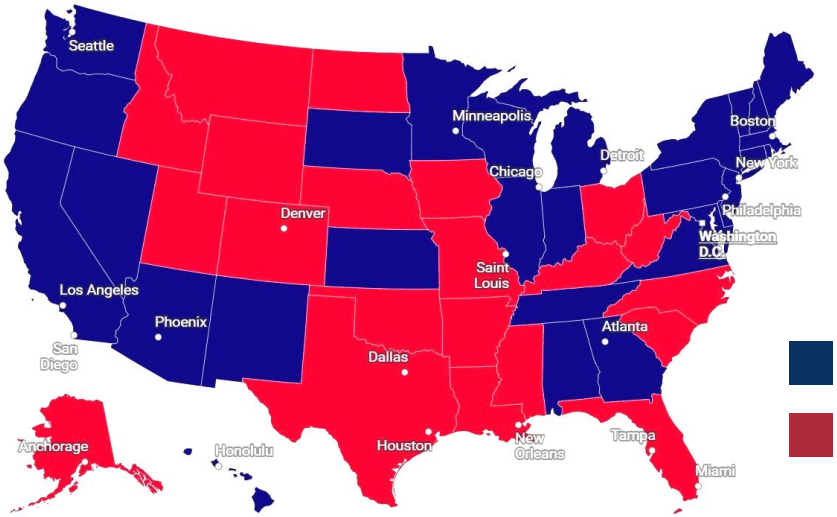


30%

P(-ve |Biden)

Comparison of our Model with Actual results

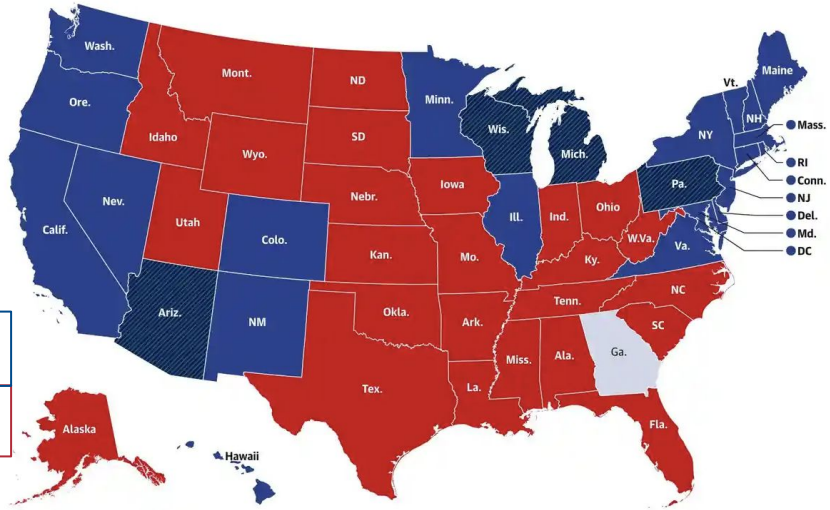
Predicted



Biden

Trump

Actual



Our model predicted the results with an accuracy of 88%

Thank You

Any Questions?

A detailed report with all reproducible code will be submitted after the presentation.